

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICANT(s): Shintaroh Hori

SERIAL NO.:

ART UNIT:

FILED: herewith

EXAMINER:

TITLE: Storage System Having Redundancy Block, And
Controller, Control Method, Program, And Storage
Medium For Storage System

ATTORNEY DOCKET NO.: JP920030055US1

Commissioner For Patents

P.O. Box 1450

Alexandria, VA 22313-1450

Transmittal Of Certified Copy

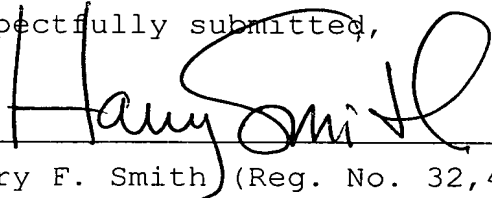
Sir:

Applicant(s) claim the benefit of the following prior foreign patent application under 35 U.S.C. §119 for the above-identified U.S. patent application:

Country: Japan
Application No.: 2003-118907
Filing Date: April 23, 2003

Attached is a certified copy of the foreign application from which priority is claimed.

Respectfully submitted,


Harry F. Smith (Reg. No. 32,493)


Date

Customer No.: 29683

Harrington & Smith, LLP

4 Research Drive

Shelton, CT 06484-6212

203-925-9400

日 本 国 特 許 庁
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出 願 年 月 日
Date of Application: 2 0 0 3 年 4 月 2 3 日

出 願 番 号
Application Number: 特 願 2 0 0 3 - 1 1 8 9 0 7

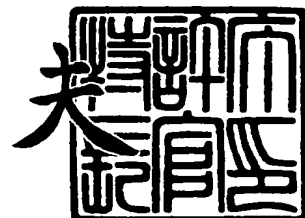
[ST. 10/C]: [J P 2 0 0 3 - 1 1 8 9 0 7]

出 願 人
Applicant(s): インターナショナル・ビジネス・マシーンズ・コーポレーション

2 0 0 3 年 8 月 2 2 日

特許庁長官
Commissioner,
Japan Patent Office

今 井 康 夫



【書類名】 特許願

【整理番号】 JP9030055

【提出日】 平成15年 4月23日

【あて先】 特許庁長官 殿

【国際特許分類】 G06F 9/45

【発明者】

 【住所又は居所】 神奈川県大和市下鶴間 1 6 2 3 番地 1 4 日本アイ・ビー・エム株式会社 大和事業所内

 【氏名】 堀 慎太郎

【特許出願人】

 【識別番号】 390009531

 【氏名又は名称】 インターナショナル・ビジネス・マシーンズ・コーポレーション

【代理人】

 【識別番号】 100086243

 【弁理士】

 【氏名又は名称】 坂口 博

【代理人】

 【識別番号】 100091568

 【弁理士】

 【氏名又は名称】 市位 嘉宏

【代理人】

 【識別番号】 100108501

 【弁理士】

 【氏名又は名称】 上野 剛史

【復代理人】

 【識別番号】 100104156

 【弁理士】

 【氏名又は名称】 龍華 明裕

【手数料の表示】

【予納台帳番号】 053394

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 9706050

【包括委任状番号】 9704733

【包括委任状番号】 0207860

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 冗長化ブロックを有する記憶システム、並びに、当該記憶システムの制御装置、制御方法、プログラム及び記録媒体

【特許請求の範囲】

【請求項 1】 複数の格納対象ブロックを含むブロックグループを、複数の記憶装置に分散して格納する記憶システムであって、

一の前記格納対象ブロックは、他の複数の前記格納対象ブロックのいずれかが故障した場合に当該格納対象ブロックを再生するための冗長データである冗長化ブロックであり、

複数の記憶装置と、

前記複数の格納対象ブロックのそれぞれと、前記複数の格納対象ブロックのいずれかを複製した複製ブロックとを、互いに異なる前記記憶装置に格納するブロック書込部と、

複製されていない前記格納対象ブロックの故障が検出された場合に、前記複数の格納対象ブロックのうち故障した前記格納対象ブロック以外のブロックに基づいて、故障した前記格納対象ブロックを再生するブロック再生部と、

再生された前記格納対象ブロックを、前記複製ブロック又は前記複製ブロックの複製元となった前記格納対象ブロックに上書きする再生ブロック上書部と

を備える記憶システム。

【請求項 2】 前記ブロック書込部は、複数の前記ブロックグループのそれぞれについて、当該ブロックグループに含まれる前記複数の格納対象ブロックのそれぞれと、前記複製ブロックとを互いに異なる前記記憶装置に格納し、

一の前記記憶装置が故障した場合に、前記ブロック再生部は、複製されていない前記格納対象ブロックが前記一の記憶装置に格納されている前記ブロックグループのそれぞれについて、前記複数の格納対象ブロックのうち前記一の記憶装置に格納されている前記格納対象ブロック以外のブロックに基づいて、故障した前記格納対象ブロックを再生し、

前記再生ブロック上書部は、複製されていない前記格納対象ブロックが前記一の記憶装置に格納されている前記ブロックグループのそれぞれについて、再生さ

れた前記格納対象ブロックを、前記複製ブロック又は前記複製ブロックの複製元となった前記格納対象ブロックに上書きする

請求項 1 記載の記憶システム。

【請求項 3】 前記ブロック書込部は、前記複数の格納対象ブロックのそれぞれと、前記複数の格納対象ブロックに含まれる前記冗長化ブロックを複製した前記複製ブロックとを、互いに異なる前記記憶装置に格納する請求項 1 記載の記憶システム。

【請求項 4】 前記複数の格納対象ブロックのうち、前記冗長化ブロック以外のブロックであるデータブロックに対して、書込データの書き込みを要求する書込要求を受信する要求受信部と、

書込対象となる前記データブロック、前記書込データ、及び元の前記冗長化ブロックに基づき、新たな前記冗長化ブロックを生成する冗長化ブロック生成部とを更に備え、

前記ブロック書込部は、書込対象となる前記データブロックに前記書込データを書き込み、元の前記冗長化ブロック及び前記複製ブロックに新たな前記冗長化ブロックを書き込む

請求項 3 記載の記憶システム。

【請求項 5】 前記ブロック書込部は、前記複数の格納対象ブロックのそれぞれと、前記複数の格納対象ブロックのうち前記冗長化ブロック以外のブロックである複数のデータブロックのいずれかを複製した前記複製ブロックとを、互いに異なる前記記憶装置に格納する請求項 1 記載の記憶システム。

【請求項 6】 前記複数の格納対象ブロックのうち、前記冗長化ブロック以外のブロックであるデータブロックに対して、書込データの書き込みを要求する書込要求を受信する要求受信部を更に備え、

前記ブロック書込部は、書込対象となる前記データブロックが前記複製元ブロックである場合に、前記複製元ブロック及び前記複製ブロックのそれぞれに前記書込データを書き込み、書込対象となる前記データブロックが前記複製元ブロックでない場合に、書込対象となる前記データブロックに前記書込データを書き込む

請求項 5 記載の記憶システム。

【請求項 7】 複数の格納対象ブロックを含むブロックグループを、複数の記憶装置に分散して格納する記憶システムの制御装置であって、

一の前記格納対象ブロックは、他の複数の前記格納対象ブロックのいずれかが故障した場合に当該格納対象ブロックを再生するための冗長データである冗長化ブロックであり、

前記複数の格納対象ブロックのそれぞれと、前記複数の格納対象ブロックのいずれかを複製した複製ブロックとを、互いに異なる前記記憶装置に格納するブロック書込部と、

複製されていない前記格納対象ブロックの故障が検出された場合に、前記複数の格納対象ブロックのうち故障した前記格納対象ブロック以外のブロックに基づいて、故障した前記格納対象ブロックを再生するブロック再生部と、

再生された前記格納対象ブロックを、前記複製ブロック又は前記複製ブロックの複製元となった前記格納対象ブロックに上書きする再生ブロック上書部とを備える制御装置。

【請求項 8】 複数の格納対象ブロックを含むブロックグループを、複数の記憶装置に分散して格納する記憶システムの制御方法であって、

一の前記格納対象ブロックは、他の複数の前記格納対象ブロックのいずれかが故障した場合に当該格納対象ブロックを再生するための冗長データである冗長化ブロックであり、

前記複数の格納対象ブロックのそれぞれと、前記複数の格納対象ブロックのいずれかを複製した複製ブロックとを、互いに異なる前記記憶装置に格納するブロック書込段階と、

複製されていない前記格納対象ブロックの故障が検出された場合に、前記複数の格納対象ブロックのうち故障した前記格納対象ブロック以外のブロックに基づいて、故障した前記格納対象ブロックを再生するブロック再生段階と、

再生された前記格納対象ブロックを、前記複製ブロック又は前記複製ブロックの複製元となった前記格納対象ブロックに上書きする再生ブロック上書段階とを備える制御方法。

【請求項 9】 複数の格納対象ブロックを含むブロックグループを、複数の記憶装置に分散して格納する記憶システムを制御するプログラムであって、

一の前記格納対象ブロックは、他の複数の前記格納対象ブロックのいずれかが故障した場合に当該格納対象ブロックを再生するための冗長データである冗長化ブロックであり、

前記記憶システムを、

前記複数の格納対象ブロックのそれぞれと、前記複数の格納対象ブロックのいずれかを複製した複製ブロックとを、互いに異なる前記記憶装置に格納するブロック書込部と、

複製されていない前記格納対象ブロックの故障が検出された場合に、前記複数の格納対象ブロックのうち故障した前記格納対象ブロック以外のブロックに基づいて、故障した前記格納対象ブロックを再生するブロック再生部と、

再生された前記格納対象ブロックを、前記複製ブロック又は前記複製ブロックの複製元となった前記格納対象ブロックに上書きする再生ブロック上書部として機能させるプログラム。

【請求項 10】 請求項 9 に記載のプログラムを記録した記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、冗長化ブロックを有する記憶システム、並びに、当該記憶システムの制御装置、制御方法、プログラム、及び記録媒体に関する。特に本発明は、複数の格納対象ブロックを含むブロックグループを、複数の記憶装置に分散して格納し、ブロックグループ中の一の格納対象ブロックが故障した場合においても、他の格納対象ブロックに基づき故障した格納対象ブロックを再生することができる冗長化ブロックを有する記憶システム、並びに、当該記憶システムの制御装置、制御方法、プログラム、及び記録媒体に関する。

【0002】

【従来の技術】

従来、ハードディスク等の記憶装置のアクセス性能を高めると共に、耐故障性

を向上するために、R A I D 技術が用いられている。例えば、R A I D 5 構成においては、複数のデータブロックと、複数のデータブロックのいずれかが故障した場合に故障したデータブロックを再生するための冗長データであるパリティブロックとにより構成される複数の格納対象ブロックを、ブロックグループとして、複数の記憶装置にブロック単位で分散して格納する。例えば、特許文献 1 は、パリティブロックを複数の記憶装置に分散して格納する R A I D 5 構成を開示する（特許文献 1 参照。）。

【 0 0 0 3 】

複数の記憶装置のうち一の記憶装置が故障した場合において、格納対象ブロックを再生し、予備のブロックに格納する技術が開示されている（特許文献 2 参照。）。また、第 1 及び第 2 のパリティブロックを設け、書き込み時に第 1 のパリティブロック又は第 2 のパリティブロックのいずれか一方を更新することにより、ブロックの書き込みに伴うパリティブロックの更新を第 1 及び第 2 のパリティブロックに分散させる技術が開示されている（特許文献 3 参照。）。

【 0 0 0 4 】

【特許文献 1】

特公平 5 - 4 7 8 5 7 号公報

【特許文献 2】

特公平 7 - 2 4 0 3 9 号公報

【特許文献 3】

特開平 9 - 3 4 6 5 1 号公報

【 0 0 0 5 】

【発明が解決しようとする課題】

特許文献 2 又は特許文献 3 に示した技術を用いた場合において、一の記憶装置が故障すると、全てのブロックグループについて失われたブロックを再生し、予備のブロックに書き込む必要が生じる。ここで、一の記憶装置が故障した状態は、更に記憶装置が故障すると記憶システムに格納された情報が失われるクリティカルな状態であるため、ブロックの再生に伴うオーバーヘッドを低減することが望ましい。

【 0 0 0 6 】

そこで本発明は、上記の課題を解決することのできる冗長化ブロックを有する記憶システム、並びに、当該記憶システムの制御装置、制御方法、プログラム、及び記録媒体を提供することを目的とする。この目的は特許請求の範囲における独立項に記載の特徴の組み合わせにより達成される。また従属項は本発明の更なる有利な具体例を規定する。

【 0 0 0 7 】**【課題を解決するための手段】**

即ち、本発明の第 1 の形態によると、複数の格納対象ブロックを含むブロックグループを、複数の記憶装置に分散して格納する記憶システムであって、一の前記格納対象ブロックは、他の複数の前記格納対象ブロックのいずれかが故障した場合に当該格納対象ブロックを再生するための冗長データである冗長化ブロックであり、複数の記憶装置と、前記複数の格納対象ブロックのそれぞれと、前記複数の格納対象ブロックのいずれかを複製した複製ブロックとを、互いに異なる前記記憶装置に格納するブロック書込部と、複製されていない前記格納対象ブロックの故障が検出された場合に、前記複数の格納対象ブロックのうち故障した前記格納対象ブロック以外のブロックに基づいて、故障した前記格納対象ブロックを再生するブロック再生部と、再生された前記格納対象ブロックを、前記複製ブロック又は前記複製ブロックの複製元となった前記格納対象ブロックに上書きする再生ブロック上書部とを備える記憶システム、並びに、当該記憶システムの制御装置、制御方法、プログラム及び記録媒体を提供する。

【 0 0 0 8 】

なお上記の発明の概要は、本発明の必要な特徴の全てを列挙したものではなく、これらの特徴群のサブコンビネーションも又発明となりうる。

【 0 0 0 9 】**【発明の実施の形態】**

以下、発明の実施の形態を通じて本発明を説明するが、以下の実施形態は特許請求の範囲にかかる発明を限定するものではなく、又実施形態の中で説明されている特徴の組み合わせの全てが発明の解決手段に必須であるとは限らない。

【0010】

図1は、本実施形態に係る記憶システム20の構成を示す。記憶システム20は、例えばRAID5等の冗長化ディスクアレイシステムであり、複数の格納対象ブロックを含むブロックグループを、複数の記憶装置30にブロック単位で分散して格納する。ここで、記憶システム20は、複数の格納対象ブロックに加え、複数の格納対象ブロックの一部をそれぞれ複製した1又は複数の複製ブロックを含めたブロックグループを、複数の記憶装置30にブロック単位で分散して格納する。

【0011】

一の記憶装置30が故障した場合、各ブロックグループ内の複製ブロックは予備のブロックとして用いられてよい。また、複製ブロックには、故障前に予め複製元の格納対象ブロックが複製されていることから、一の記憶装置に複製元のブロック又は複製ブロックが格納されているブロックグループについては、故障したブロックの再生が不要となり、クリティカルな状態におけるデータ再生処理のオーバーヘッドを低減することができる。

【0012】

記憶システム20は、情報処理装置10に接続され、情報処理装置10からの要求に応じてデータの書き込み及び／又は読み出しを行なう。記憶システム20は、複数の記憶装置30と、制御部40とを備える。

【0013】

複数の記憶装置30のそれぞれは、例えばハードディスク等の、データをブロック単位で格納する記憶装置である。複数の記憶装置30は、複数の格納対象ブロックと、複数の格納対象ブロックの一部をそれぞれ複製した1又は複数の複製ブロックとを含むブロックグループを、ブロック単位で分散して格納する。

【0014】

複数の格納対象ブロックの一部、例えば本実施形態においては一の格納対象ブロックは、他の複数の格納対象ブロックのいずれかが故障した場合に当該格納対象ブロックを再生するための冗長データである冗長化ブロックである。また、複数の格納対象ブロックのうち、冗長化ブロック以外のブロックは、情報処理装置

1 0 が使用するデータを格納するデータブロックである。

【0 0 1 5】

以下本実施形態においては、4 つのデータブロックと、1 つの冗長化ブロックと、1 つの複製ブロックとを一のブロックグループとして、記憶装置 3 0 a ~ f にブロック単位で分散して格納する場合を例として説明する。

【0 0 1 6】

制御部 4 0 は、複数の記憶装置 3 0 を制御する。制御部 4 0 は、要求受信部 4 5 と、応答送信部 5 0 と、ブロック書込部 5 5 と、ブロック読出部 6 0 と、故障検出部 6 5 と、ブロック再生部 7 0 と、再生ブロック上書部 7 5 と、冗長化ブロック生成部 8 0 とを有する。

【0 0 1 7】

要求受信部 4 5 は、情報処理装置 1 0 からデータブロックの読出要求又は書込要求等のコマンドを受信する。書込要求の場合、要求受信部 4 5 は、書込要求のコマンドと共に、書込データを受信する。応答送信部 5 0 は、コマンドに対する処理結果等の応答を情報処理装置 1 0 に送信する。ここで、応答送信部 5 0 は、読出要求に対する処理結果として、読み出されたデータブロックを情報処理装置 1 0 に送信してよい。

【0 0 1 8】

ブロック書込部 5 5 は、ブロックグループ毎に、複数の格納対象ブロックのそれぞれと、複数の格納対象ブロックのいずれかを複製した複製ブロックとを、互いに異なる記憶装置 3 0 に格納する。ブロック読出部 6 0 は、情報処理装置 1 0 からの要求を受けた場合、又は故障した記憶装置 3 0 を交換した場合において新たな記憶装置 3 0 の再構築を行なう場合に、記憶装置 3 0 に格納された読出対象のブロックを読み出す。

【0 0 1 9】

故障検出部 6 5 は、記憶装置 3 0 に格納されたブロックがブロック読出部 6 0 により読み出せない場合や、読み出されたブロックにエラーが存在する場合等に、当該ブロックが故障したことを検出する。また、故障検出部 6 5 は、複数の記憶装置 3 0 のうち、いずれか一の記憶装置 3 0 が故障したことを検出する。この

場合、故障検出部 65 は、当該一の記憶装置 30 に格納された全てのブロックが故障したものとして検出する。

【0020】

ブロック再生部 70 は、故障検出部 65 により複製されていない格納対象ブロックの故障が検出された場合に、故障した格納対象ブロックが含まれるブロックグループが有する複数の格納対象ブロックのうち、故障した当該格納対象ブロック以外のブロックをブロック読出部 60 を介して読み出す。そして、ブロック再生部 70 は、これらのブロックに基づいて故障した格納対象ブロックを再生する。再生ブロック上書部 75 は、故障検出部 65 により再生された格納対象ブロックを、当該ブロックグループ内の複製ブロック、又は、複製ブロックの複製元となった格納対象ブロックである複製元ブロックに上書きする。

【0021】

冗長化ブロック生成部 80 は、要求受信部 45 が情報処理装置 10 からの書込要求を受信した場合に、ブロック読出部 60 を介して読み出した書込対象となるデータブロック及び元の冗長化ブロックと、書込データとに基づき、新たな冗長化ブロックを生成する。ブロック読出部 60 が生成した新たな冗長化ブロックは、ブロック書込部 55 により記憶装置 30 に書き込まれる。

【0022】

以上に示した記憶システム 20 によれば、各ブロックグループについて、複数の格納対象ブロックのそれぞれと、複数の格納対象ブロックのいずれかを複製した複製ブロックとを、互いに異なる記憶装置 30 に格納する。そして、ブロックの故障や記憶装置 30 の故障により一のブロックが故障した場合、故障したブロックが複製ブロック又は複製元ブロックであった場合にはブロックの再生が不要となる。これにより、ブロックの再生に伴うオーバーヘッドを低減することができる。

【0023】

図 2 は、本実施形態に係る記憶装置 30 に格納されるブロックの配置の一例を示す。ブロック書込部 55 は、記憶システム 20 に格納される複数のブロックグループのそれぞれについて、当該ブロックグループに含まれる複数の格納対象ブ

ロックのそれぞれと、複数の格納対象ブロックのいずれかを複製した複製ブロックとを、互いに異なる記憶装置 30 に格納する。本例において、ブロック書込部 55 は、複数のブロックグループのそれぞれについて、当該ブロックグループの複数の格納対象ブロックに含まれる冗長化ブロックを複製し、複製ブロックとして用いる。

【0024】

例えば、ストライプ 1 と示したブロックグループは、図中 DB 1 a、DB 1 b、DB 1 c、及び DB 1 d と示した 4 つのデータブロックと、図中 PB 1 と示した冗長化ブロックと、図中 PB 1' と示した冗長化ブロックの複製ブロックとを、複数の記憶装置 30 に分散して格納する。本例において、冗長化ブロック PB 1 は、複製ブロック PB 1' の複製元となる複製元ブロックである。また、ブロック書込部 55 は、複数のブロックグループについて、複製元ブロック PB 1、PB 2、…、PB 6 及び複製ブロック PB 1'、PB 2'、…、PB 6' を、複数の記憶装置 30 に分散して格納されるように、インターリーブして格納される。

【0025】

図 3 は、本実施形態に係る記憶システム 20 における書込処理の流れを示す。

まず、要求受信部 45 は、ブロックグループに含まれる複数の格納対象ブロックのうち、冗長化ブロック以外のブロックであるデータブロックに対して、書込データの書き込みを要求する書込要求を受信する (S300)。書込要求を受信すると、ブロック読出部 60 は、書込対象となるデータブロック及び元の冗長化ブロックを読み出す (S310、S320)。

【0026】

次に、冗長化ブロック生成部 80 は、ブロック読出部 60 により読み出された書込対象となるデータブロック及び元の冗長化ブロックと、要求受信部 45 が受信した書込データとに基づき、新たな冗長化ブロックを生成する (S330)。より具体的には、記憶システム 20 が RAID 5 構成を採る場合、冗長化ブロック生成部 80 は、書込対象となるデータブロック、元の冗長化ブロック、及び書込データをビット毎に排他的論理和演算することにより、新たな冗長化ブロック

を生成する。次に、ブロック書込部 55 は、書込対象となるデータブロックに書き込みデータを書き込み（S340）、元の冗長化ブロック及び複製ブロックに新たな冗長化ブロックを書き込む（S350）。

【0027】

以上に示した書込処理により、制御部 40 は、複数のブロックグループのそれぞれについて、複製ブロックと複製元ブロックである冗長化ブロックとを同一に保つことができる。ここで、ブロック書込部 55 は、新たな冗長化ブロックを、冗長化ブロック及び複製ブロックに並列に書き込むことにより、2つの記憶装置 30 に新たな冗長化ブロックを書き込む処理時間を低減することができる。

【0028】

図4は、図2において、記憶装置 30b が故障した状態を示す。図4に示す様に、一の記憶装置 30 が故障すると、当該記憶装置 30 に格納された全てのブロックが読み出し不可能となる。このため、故障検出部 65 は、当該記憶装置 30 に格納された全てのブロックが故障したものとして検出する。

ここで、図2に示す様に、ブロック書込部 55 は、複数の複製元ブロック及び複製ブロックをインターリーブして格納する。このため、記憶装置 30b の故障に伴い、ストライプ 1、2、3、及び 6 についてはデータブロックが故障し、ストライプ 4 については複製元ブロックが故障し、ストライプ 5 については、複製ブロックが故障する。

【0029】

図5は、本実施形態に係る記憶システム 20 における再構築処理の流れを、記憶装置 30 が故障した場合について示す。

まず、故障検出部 65 は、記憶装置 30 の故障を検出する（S500）。故障を検出した場合、制御部 40 は、記憶システム 20 が格納する複数のブロックグループのそれぞれについて、S520、S530、S540、及び S550 からなるループ処理を行なう（S510、S560）。

【0030】

ループ処理により再構築の対象となったブロックグループにおいて、複製されていない格納対象ブロックが故障したものとして検出された場合（S520のY

ES)、ブロック再生部70は、複数の格納対象ブロックのうち故障した格納対象ブロック以外のブロックをブロック読出部60を介して記憶装置30からそれぞれ読み出し、読み出したブロックに基づいて故障した格納対象ブロックを再生する(S540)。次に、再生ブロック上書部75は、再生された格納対象ブロックを、複製ブロック又は複製元ブロックに上書きする(S550)。

【0031】

以上の処理により、一の記憶装置30が故障した場合に、ブロック再生部70は、複製されていない格納対象ブロックが当該記憶装置30に格納されているブロックグループのそれぞれについて、複数の格納対象ブロックのうち当該記憶装置30に格納されている格納対象ブロック以外のブロックに基づいて、故障した格納対象ブロックを再生する。そして、再生ブロック上書部75は、複製されていない格納対象ブロックが当該記憶装置30に格納されているブロックグループのそれぞれについて、再生された格納対象ブロックを、複製ブロック又は複製元の格納対象ブロックに上書きすることができる。

【0032】

ここで、一の記憶装置30が故障した場合において記憶システム20の全ブロックグループを再構築する処理のオーバヘッドは、次の様に見積もることができる。本実施形態に係る記憶システム20においては、全ブロックグループのうち、複製されていない格納対象ブロックが故障した記憶装置30に格納されているブロックグループのそれぞれについて、故障した格納対象ブロックの再生を行なう。一方、複製元ブロック又は複製ブロックの一方が故障した記憶装置30に格納されているブロックグループについては、故障したブロックの再生が不要となる。従って、複製ブロック及び複製元ブロックが複数の記憶装置30に均一にインターリーブされている場合、記憶装置の数をN、ブロックグループ当たりの複製ブロックの数を1とすると、全ブロックグループのうち、再生を行なうブロックグループの比率は $(N-2)/N$ となる。

【0033】

一方、複製ブロックを持たず、予備のブロックを有する特許文献2の記憶システムにおいては、全ブロックグループのうち、故障した記憶装置30に予備のブ

ロックが格納されているブロックグループについて、故障したブロックの再生が不要となる。従って、全ブロックグループのうち、再生を行なうブロックグループの比率は $(N-1)/N$ となる。

【0034】

以上により、上記の例において、本実施形態に係る記憶システム 20 によれば、予備のブロックを用いる記憶システムと比較し再構築処理において再生を行なうブロックグループの数を $(N-2)/(N-1)$ に削減することができる。ここで、故障した記憶装置 30 を再構築するために複数の記憶装置 30 に発行される I/O 命令の数は、再生を行なうブロックグループの数と比例する。このため、記憶システムが単位時間に処理可能な I/O 命令数が同一である場合、上記の例においては、本実施形態に係る記憶システム 20 により、再構築処理の性能を $(N-1)/(N-2)$ 倍に高めることができる。

【0035】

図 6 は、図 4 において、ブロックを再構築した後のブロック配置の一例を示す。図中 () で示したブロックは、再構築により再生されたブロックである。

記憶装置 30 b が故障した場合、ブロック再生部 70 は、複製されていない格納対象ブロックが記憶装置 30 b に格納されているストライプ 1、2、3、及び 6 のそれぞれについて、複数の格納対象ブロックのうち記憶装置 30 b に格納されているブロック以外のブロックに基づいて、故障した格納ブロックを再生する。そして、再生ブロック上書部 75 は、再生された格納対象ブロックを、複製ブロック又は複製元ブロックに上書きする。

【0036】

すなわち例えば、ストライプ 1 においては、ブロック再生部 70 は、DB 1 a、DB 1 c、DB 1 d、及び、複製ブロック PB 1 に基づいて、故障した DB 1 b を再生する。そして、再生ブロック上書部 75 は、再生された DB 1 b を、記憶装置 30 e 内の複製元ブロック PB 1 又は記憶装置 30 f 内の複製ブロック PB 1' に上書きする。なお、図 6 においては、再生された DB 1 b を、複製ブロック PB 1' に上書きした状態を示す。

【0037】

ブロックを再構築することにより、制御部 40 は、複数のブロックグループのそれぞれについて、全ての格納対象ブロックが故障していない記憶装置 30 に格納された状態とすることができる。ここで、故障した記憶装置 30 に複製元ブロック又は複製ブロックが格納されているブロックグループについては、故障したブロックの再生が不要となるため、データ再生処理のオーバーヘッドを低減することができる。

【0038】

図 7 は、図 6 において、記憶装置 30 b を交換した後ブロックを再書き込みした状態の一例を示す。

故障した記憶装置 30 が新たな記憶装置 30 に交換されると、制御部 40 は、複数のブロックグループのそれぞれについて、再生された格納対象ブロック、又は、故障により失われた複製ブロック若しくは複製元ブロックをブロック読出部 60 により読み出し、ブロック書込部 55 により記憶装置 30 b に書き戻す。この処理により書き戻されたブロックを、図中 [] により示す。

【0039】

この後制御部 40 は、複数のブロックグループのうち、再生された格納対象ブロックを有するブロックグループのそれぞれについて、ブロック読出部 60 により冗長化ブロックを読み出して、ブロック書込部 55 により再生された格納対象ブロックに上書きし、図 2 に示したブロック配置に復旧させる。

【0040】

以上に示した様に、故障した記憶装置 30 が新たな記憶装置 30 に交換された場合に、制御部 40 は、まず全てのブロックグループのそれぞれについて、再生された格納対象ブロック、又は、故障により失われた複製ブロック若しくは複製元ブロックを記憶装置 30 b に書き戻す。この状態において一の記憶装置 30 が別途故障した場合においても、複数のブロックグループのそれぞれについて、最大で 1 の格納対象ブロックの内容が失われるのみであり、他の記憶装置 30 が格納するブロックに基づいて故障した記憶装置 30 に格納されたブロックを再生可能である。従って、新たな記憶装置 30 に交換された場合に、より早くクリティカルな状態を脱することができる。

【0041】

図8は、本実施形態の変形例に係る記憶装置30に格納されるブロックの配置の一例を示す。本変形例において、ブロック書込部55は、記憶システム20に格納される複数のブロックグループのそれぞれについて、当該ブロックグループに含まれる複数の格納部ロックのそれぞれと、複数の格納対象ブロックのうち冗長化ブロック以外のブロックである複数のデータブロックのいずれかを複製した複製ブロックとを、互いに異なる記憶装置30に格納する。

【0042】

例えば、ストライプ1と示したブロックグループは、図中DB1a、DB1b、DB1c、及びDB1dと示した4つのデータブロックと、図中PB1と示した冗長化ブロックと、図中DB1a' と示したデータブロックDB1aの複製ブロックとを、複数の記憶装置30に分散して格納する。

【0043】

図9は、本実施形態の変形例に係る記憶システム20における書込処理の流れを示す。

S900、S910、S920、及びS930は、図3のS300、S310、S320、及びS330とそれぞれ同様の処理であるため説明を省略する。

【0044】

書込対象となるデータブロックが複製元ブロックである場合（S940のYes）、ブロック書込部55は、複製元ブロック及び複製ブロックのそれぞれに書込データを書き込む（S950）。一方、書込対象となるデータブロックが複製元ブロックでない場合（S940のNo）、ブロック書込部55は、書込対象となるデータブロックに書込データを書き込む（S960）。そして、ブロック書込部55は、元の冗長化ブロックに新たな冗長化ブロックを書き込む（S970）。

【0045】

以上に示した記憶システム20によれば、書込対象となるデータブロックが複製元ブロックでない場合に、書込データを当該データブロックに書き込むと共に、冗長化ブロックを更新するのみで、複製元ブロックと複製ブロックとを同一に

保つことができる。これにより、冗長化ブロックを複製元ブロックとした場合と比較し、複数の記憶装置 30 に対するブロック書込部 55 の書き込み回数を低減することができる。

【0046】

図 10 は、図 8 において記憶装置 30 b が故障し、ブロックを再構築した後のブロック配置の一例を示す。図中 () で示したブロックは、再構築により再生されたブロックである。

本変形例において一の記憶装置 30 が故障した場合、制御部 40 は、図 5 に示した再構築処理により、複数の格納対象ブロックのうち故障していない格納対象ブロックに基づいて、故障した格納対象ブロックを再生し、複製元ブロック又は複製ブロックに上書きする。例えば記憶装置 30 b が故障した場合、再生ブロック上書部 75 は、図 10 に示す様に、複製されていない格納対象ブロックが記憶装置 30 b に格納されているストライプ 1 から 4 のそれぞれについて、複製されていないデータブロック又は冗長化ブロックを再生する。そして、再生ブロック上書部 75 は、再生されたデータブロック又は冗長化ブロックを、複製元ブロック又は複製ブロックの一方（本例においては複製ブロック）に上書きする。

【0047】

本変形例によれば、一の記憶装置 30 が故障した場合において、記憶システム 20 は、冗長化ブロックを複製する場合と同様にして再構築することにより、全ての格納対象ブロックが故障していない記憶装置 30 に格納された状態とすることができる。ここで、故障した記憶装置 30 に複製元ブロック又は複製ブロックが格納されているブロックグループについては、故障したブロックの再生が不要となるため、データ再生処理のオーバーヘッドを低減することができる。

【0048】

図 11 は、本実施形態に係る情報処理装置 10 のハードウェア構成の一例を示す。本実施形態に係る情報処理装置 10 は、ホスト・コントローラ 1182 により相互に接続される CPU 1100、RAM 1120、グラフィック・コントローラ 1175、及び表示装置 1180 を有する CPU 周辺部と、入出力コントローラ 1184 によりホスト・コントローラ 1182 に接続される通信インターフ

エイス 1130、ハードディスク・ドライブ 1140、及び CD-ROM ドライブ 1160 を有する入出力部と、入出力コントローラ 1184 に接続される ROM 1110、フレキシブルディスク・ドライブ 1150、及び入出力チップ 1170 を有するレガシー入出力部とを備える。

【0049】

ホスト・コントローラ 1182 は、RAM 1120 と、高い転送レートで RAM 1120 をアクセスする CPU 1100 及びグラフィック・コントローラ 1175 とを接続する。CPU 1100 は、ROM 1110 及び RAM 1120 に格納されたプログラムに基づいて動作し、各部の制御を行う。グラフィック・コントローラ 1175 は、CPU 1100 等が RAM 1120 内に設けたフレーム・バッファ上に生成する画像データを取得し、表示装置 1180 上に表示させる。これに代えて、グラフィック・コントローラ 1175 は、CPU 1100 等が生成する画像データを格納するフレーム・バッファを、内部に含んでもよい。

【0050】

入出力コントローラ 1184 は、ホスト・コントローラ 1182 と、比較的高速な入出力装置である通信インターフェイス 1130、ハードディスク・ドライブ 1140、CD-ROM ドライブ 1160 を接続する。通信インターフェイス 1130 は、ネットワークを介して記憶システム 20 や他の装置と通信する。ハードディスク・ドライブ 1140 は、情報処理装置 10 内の CPU 1100 が使用するプログラム及びデータを格納する。CD-ROM ドライブ 1160 は、CD-ROM 1195 からプログラム又はデータを読み取り、入出力コントローラ 1184 及び通信インターフェイス 1130 を介して記憶システム 20 に提供する。

【0051】

また、入出力コントローラ 1184 には、ROM 1110 と、フレキシブルディスク・ドライブ 1150 や入出力チップ 1170 等の比較的低速な入出力装置とが接続される。ROM 1110 は、情報処理装置 1000 が起動時に実行するブート・プログラムや、情報処理装置 10 のハードウェアに依存するプログラム等を格納する。フレキシブルディスク・ドライブ 1150 は、フレキシブルディ

スク 1 1 9 0 からプログラム又はデータを読み取り、入出力コントローラ 1 1 8 4 及び通信インターフェイス 1 1 3 0 を介して記憶システム 2 0 に提供する。入出力チップ 1 1 7 0 は、フレキシブルディスク 1 1 9 0 や、例えばパラレル・ポート、シリアル・ポート、キーボード・ポート、マウス・ポート等を介して各種の入出力装置を接続する。

【 0 0 5 2 】

R A M 1 1 2 0 を介して記憶システム 2 0 に提供されるプログラムは、フレキシブルディスク 1 1 9 0、C D - R O M 1 1 9 5、又は I C カード等の記録媒体に格納されて利用者によって提供される。プログラムは、記録媒体から読み出され、入出力コントローラ 1 1 8 4 及び通信インターフェイス 1 1 3 0 を介して記憶システム 2 0 内の制御部 4 0 にインストールされ、制御部 4 0 において実行される。

【 0 0 5 3 】

記憶システム 2 0 内の制御部 4 0 にインストールされて実行されるプログラムは、記憶システム 2 0 内の制御部 4 0 を、要求受信モジュールと、応答送信モジュールと、ブロック書込モジュールと、ブロック読出モジュールと、故障検出モジュールと、ブロック再生モジュールと、再生ブロック上書モジュールと、冗長化ブロック生成モジュールとを、要求受信部 4 5、応答送信部 5 0、ブロック書込部 5 5、ブロック読出部 6 0、故障検出部 6 5、ブロック再生部 7 0、再生ブロック上書部 7 5 及び冗長化ブロック生成部 8 0 としてそれぞれ機能させる。

【 0 0 5 4 】

以上に示したプログラム又はモジュールは、外部の記憶媒体に格納されてもよい。記憶媒体としては、フレキシブルディスク 1 1 9 0、C D - R O M 1 1 9 5 の他に、D V D や P D 等の光学記録媒体、M D 等の光磁気記録媒体、テープ媒体、I C カード等の半導体メモリ等を用いることができる。また、専用通信ネットワークやインターネットに接続されたサーバシステムに設けたハードディスク又は R A M 等の記憶装置を記録媒体として使用し、ネットワークを介してプログラムを記憶システム 2 0 に提供してもよい。

【 0 0 5 5 】

以上、本発明を実施形態を用いて説明したが、本発明の技術的範囲は上記実施形態に記載の範囲には限定されない。上記実施形態に、多様な変更または改良を加えることができる。そのような変更または改良を加えた形態も本発明の技術的範囲に含まれ得ることが、特許請求の範囲の記載から明らかである。

【0056】

以上に説明した実施形態によれば、以下の各項目に示す冗長化ブロックを有する記憶システム、並びに、当該記憶システムの制御装置、制御方法、プログラム、及び記録媒体が実現される。

【0057】

(項目1) 複数の格納対象ブロックを含むブロックグループを、複数の記憶装置に分散して格納する記憶システムであって、一の前記格納対象ブロックは、他の複数の前記格納対象ブロックのいずれかが故障した場合に当該格納対象ブロックを再生するための冗長データである冗長化ブロックであり、複数の記憶装置と、前記複数の格納対象ブロックのそれぞれと、前記複数の格納対象ブロックのいずれかを複製した複製ブロックとを、互いに異なる前記記憶装置に格納するブロック書込部と、複製されていない前記格納対象ブロックの故障が検出された場合に、前記複数の格納対象ブロックのうち故障した前記格納対象ブロック以外のブロックに基づいて、故障した前記格納対象ブロックを再生するブロック再生部と、再生された前記格納対象ブロックを、前記複製ブロック又は前記複製ブロックの複製元となった前記格納対象ブロックに上書きする再生ブロック上書部とを備える記憶システム。

【0058】

(項目2) 前記ブロック書込部は、複数の前記ブロックグループのそれぞれについて、当該ブロックグループに含まれる前記複数の格納対象ブロックのそれぞれと、前記複製ブロックとを互いに異なる前記記憶装置に格納し、一の前記記憶装置が故障した場合に、前記ブロック再生部は、複製されていない前記格納対象ブロックが前記一の記憶装置に格納されている前記ブロックグループのそれぞれについて、前記複数の格納対象ブロックのうち前記一の記憶装置に格納されている前記格納対象ブロック以外のブロックに基づいて、故障した前記格納対象ブ

ックを再生し、前記再生ブロック上書部は、複製されていない前記格納対象ブロックが前記一の記憶装置に格納されている前記ブロックグループのそれぞれについて、再生された前記格納対象ブロックを、前記複製ブロック又は前記複製ブロックの複製元となった前記格納対象ブロックに上書きする項目 1 記載の記憶システム。

【 0 0 5 9 】

(項目 3) 前記ブロック書込部は、前記複数の格納対象ブロックのそれぞれと、前記複数の格納対象ブロックに含まれる前記冗長化ブロックを複製した前記複製ブロックとを、互いに異なる前記記憶装置に格納する項目 1 記載の記憶システム。

(項目 4) 前記複数の格納対象ブロックのうち、前記冗長化ブロック以外のブロックであるデータブロックに対して、書込データの書き込みを要求する書込要求を受信する要求受信部と、書込対象となる前記データブロック、前記書込データ、及び元の前記冗長化ブロックに基づき、新たな前記冗長化ブロックを生成する冗長化ブロック生成部とを更に備え、前記ブロック書込部は、書込対象となる前記データブロックに前記書込データを書き込み、元の前記冗長化ブロック及び前記複製ブロックに新たな前記冗長化ブロックを書き込む項目 3 記載の記憶システム。

【 0 0 6 0 】

(項目 5) 前記ブロック書込部は、前記複数の格納対象ブロックのそれぞれと、前記複数の格納対象ブロックのうち前記冗長化ブロック以外のブロックである複数のデータブロックのいずれかを複製した前記複製ブロックとを、互いに異なる前記記憶装置に格納する項目 1 記載の記憶システム。

(項目 6) 前記複数の格納対象ブロックのうち、前記冗長化ブロック以外のブロックであるデータブロックに対して、書込データの書き込みを要求する書込要求を受信する要求受信部を更に備え、前記ブロック書込部は、書込対象となる前記データブロックが前記複製元ブロックである場合に、前記複製元ブロック及び前記複製ブロックのそれぞれに前記書込データを書き込み、書込対象となる前記データブロックが前記複製元ブロックでない場合に、書込対象となる前記データ

ブロックに前記書込データを書き込む項目 5 記載の記憶システム。

【0061】

(項目 7) 複数の格納対象ブロックを含むブロックグループを、複数の記憶装置に分散して格納する記憶システムの制御装置であって、一の前記格納対象ブロックは、他の複数の前記格納対象ブロックのいずれかが故障した場合に当該格納対象ブロックを再生するための冗長データである冗長化ブロックであり、前記複数の格納対象ブロックのそれぞれと、前記複数の格納対象ブロックのいずれかを複製した複製ブロックとを、互いに異なる前記記憶装置に格納するブロック書込部と、複製されていない前記格納対象ブロックの故障が検出された場合に、前記複数の格納対象ブロックのうち故障した前記格納対象ブロック以外のブロックに基づいて、故障した前記格納対象ブロックを再生するブロック再生部と、再生された前記格納対象ブロックを、前記複製ブロック又は前記複製ブロックの複製元となった前記格納対象ブロックに上書きする再生ブロック上書部とを備える制御装置。

【0062】

(項目 8) 複数の格納対象ブロックを含むブロックグループを、複数の記憶装置に分散して格納する記憶システムの制御方法であって、一の前記格納対象ブロックは、他の複数の前記格納対象ブロックのいずれかが故障した場合に当該格納対象ブロックを再生するための冗長データである冗長化ブロックであり、前記複数の格納対象ブロックのそれぞれと、前記複数の格納対象ブロックのいずれかを複製した複製ブロックとを、互いに異なる前記記憶装置に格納するブロック書込段階と、複製されていない前記格納対象ブロックの故障が検出された場合に、前記複数の格納対象ブロックのうち故障した前記格納対象ブロック以外のブロックに基づいて、故障した前記格納対象ブロックを再生するブロック再生段階と、再生された前記格納対象ブロックを、前記複製ブロック又は前記複製ブロックの複製元となった前記格納対象ブロックに上書きする再生ブロック上書段階とを備える制御方法。

【0063】

(項目 9) 複数の格納対象ブロックを含むブロックグループを、複数の記憶装

置に分散して格納する記憶システムを制御するプログラムであって、一の前記格納対象ブロックは、他の複数の前記格納対象ブロックのいずれかが故障した場合に当該格納対象ブロックを再生するための冗長データである冗長化ブロックであり、前記記憶システムを、前記複数の格納対象ブロックのそれぞれと、前記複数の格納対象ブロックのいずれかを複製した複製ブロックとを、互いに異なる前記記憶装置に格納するブロック書込部と、複製されていない前記格納対象ブロックの故障が検出された場合に、前記複数の格納対象ブロックのうち故障した前記格納対象ブロック以外のブロックに基づいて、故障した前記格納対象ブロックを再生するブロック再生部と、再生された前記格納対象ブロックを、前記複製ブロック又は前記複製ブロックの複製元となった前記格納対象ブロックに上書きする再生ブロック上書部として機能させるプログラム。

(項目 1 0) 項目 9 に記載のプログラムを記録した記録媒体。

【 0 0 6 4 】

【発明の効果】

上記説明から明らかなように、本発明によれば、複数の格納対象ブロックからなるブロックグループの複数の記憶装置に分散して格納する記憶システムにおいて、一の記憶装置が故障した場合におけるブロック再生処理のオーバーヘッドを低減することができる。

【図面の簡単な説明】

【図 1】 本発明の実施形態に係る記憶システム 2 0 の構成を示す。

【図 2】 本発明の実施形態に係る記憶装置 3 0 に格納されるブロックの配置の一例を示す。

【図 3】 本発明の実施形態に係る記憶システム 2 0 における書込処理の流れを示す。

【図 4】 図 2 において、記憶装置 3 0 b が故障した状態を示す。

【図 5】 本発明の実施形態に係る記憶システム 2 0 における再構築処理の流れを示す。

【図 6】 図 4 において、ブロックを再構築した後のブロック配置の一例を示す。

【図 7】 図 6 において、記憶装置 3 0 b を交換した後ブロックを再書込みした状態を示す。

【図 8】 本発明の実施形態の変形例に係る記憶装置 3 0 に格納されるブロックの配置の一例を示す。

【図 9】 本発明の実施形態の変形例に係る記憶システム 2 0 における書込処理の流れを示す。

【図 1 0】 図 8 において記憶装置 3 0 b が故障し、ブロックを再構築した後のブロック配置の一例を示す。

【図 1 1】 本発明の実施形態に係る情報処理装置 1 0 のハードウェア構成の一例を示す。

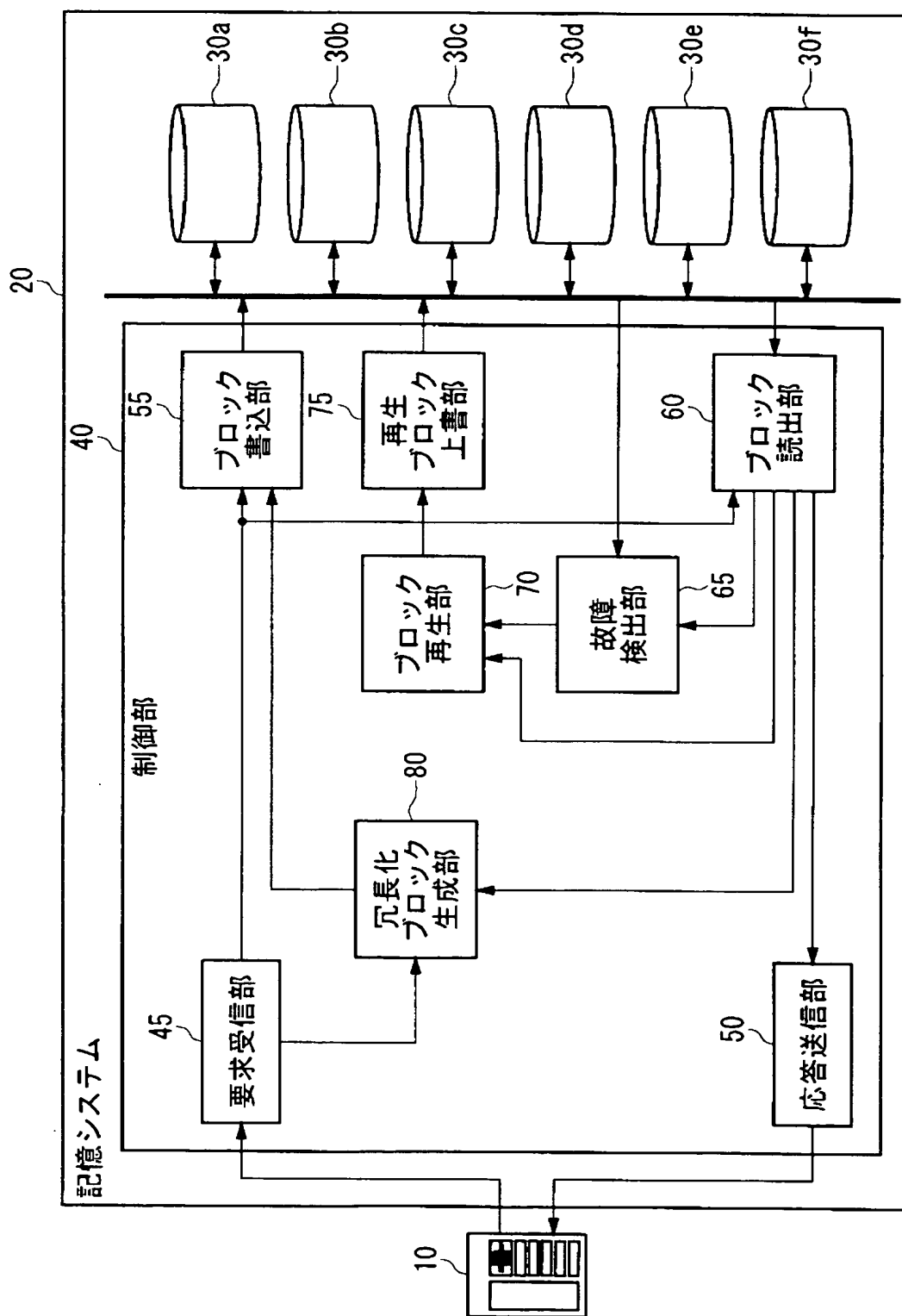
【符号の説明】

- 1 0 情報処理装置
- 2 0 記憶システム
- 3 0 a ~ f 記憶装置
- 4 0 制御部
- 4 5 要求受信部
- 5 0 応答送信部
- 5 5 ブロック書込部
- 6 0 ブロック読出部
- 6 5 故障検出部
- 7 0 ブロック再生部
- 7 5 再生ブロック上書部
- 8 0 冗長化ブロック生成部
- 1 1 0 0 C P U
- 1 1 1 0 R O M
- 1 1 2 0 R A M
- 1 1 3 0 通信インターフェイス
- 1 1 4 0 ハードディスク・ドライブ
- 1 1 5 0 フレキシブルディスク・ドライブ

1 1 6 0 C D - R O M ド ラ イ ブ
1 1 7 0 入出力チップ
1 1 7 5 グラフィック・コントローラ
1 1 8 0 表示装置
1 1 8 2 ホスト・コントローラ
1 1 8 4 入出力コントローラ
1 1 9 0 フレキシブルディスク
1 1 9 5 C D - R O M

【書類名】 図面

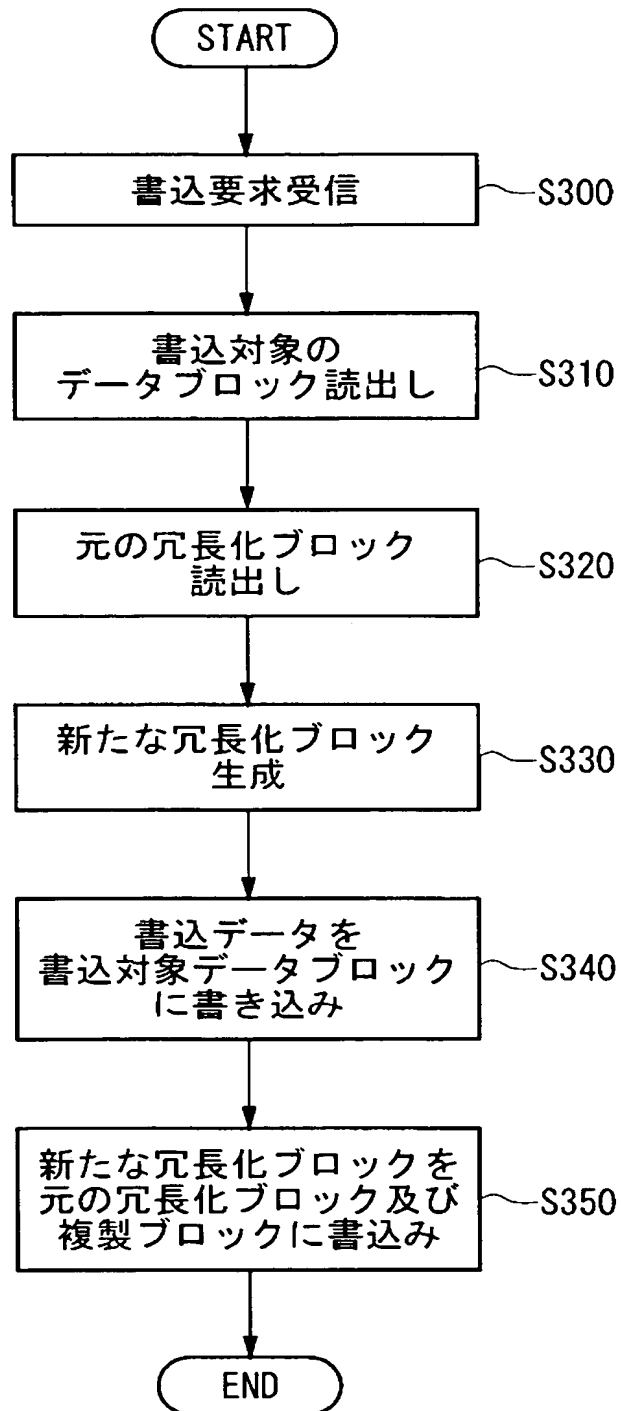
【図 1】



【図 2】

	記憶装置 30a	記憶装置 30b	記憶装置 30c	記憶装置 30d	記憶装置 30e	記憶装置 30f
ストライプ 1	DB1a	DB1b	DB1c	DB1d	PB1	PB1'
ストライプ 2	DB2b	DB2c	DB2d	PB2	PB2'	DB2a
ストライプ 3	DB3c	DB3d	PB3	PB3'	DB3a	DB3b
ストライプ 4	DB4d	PB4	PB4'	DB4a	DB4b	DB4c
ストライプ 5	PB5	PB5'	DB5a	DB5b	DB5c	DB5d
ストライプ 6	PB6'	DB6a	DB6b	DB6c	DB6d	PB6

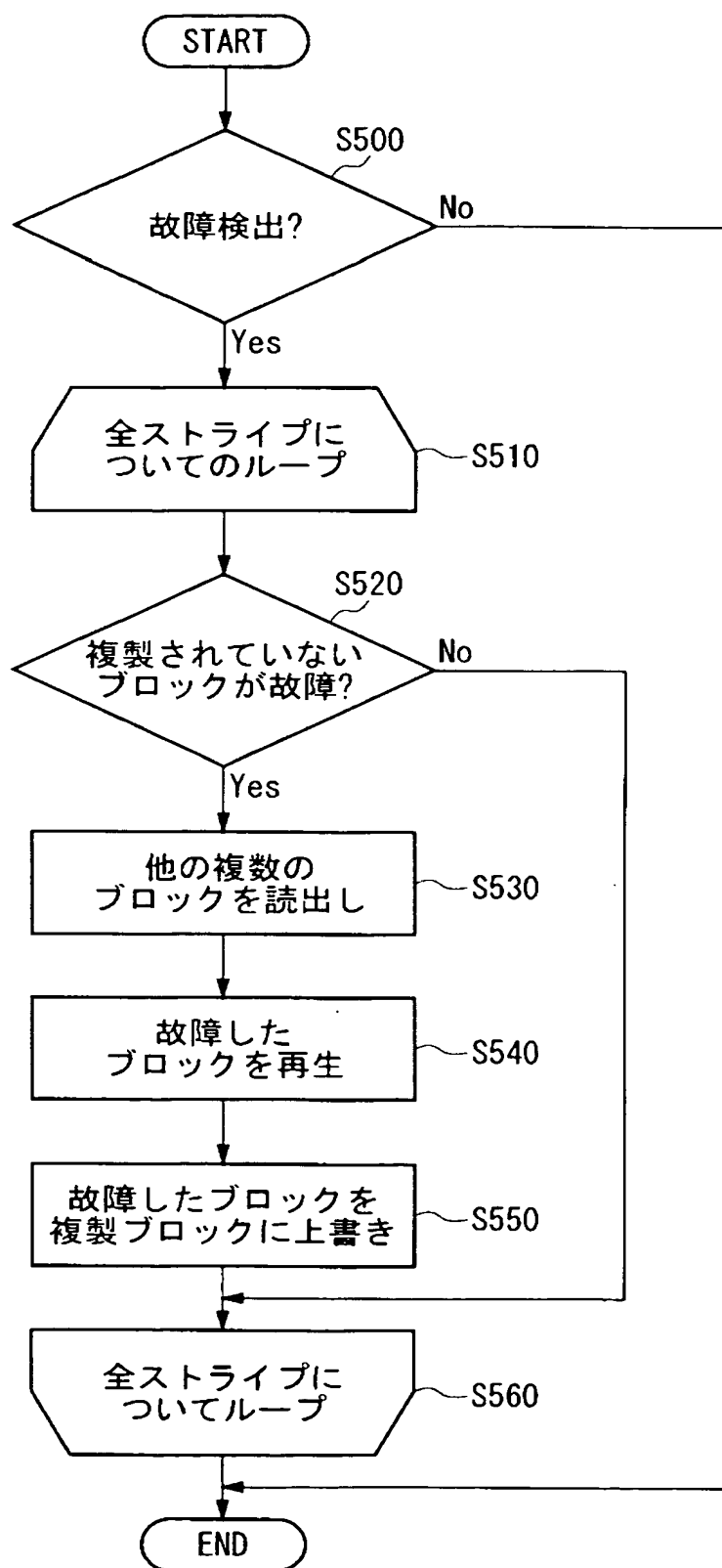
【図 3】



【図 4】

	記憶装置 30a	記憶装置 30b	記憶装置 30c	記憶装置 30d	記憶装置 30e	記憶装置 30f
ストライプ 1	DB1a	DB1b	DB1c	DB1d	PB1	PB1'
ストライプ 2	DB2b	DB2c	DB2d	PB2	PB2'	DB2a
ストライプ 3	DB3c	DB3d	PB3	PB3'	DB3a	DB3b
ストライプ 4	DB4d	PB4	PB4'	DB4a	DB4b	DB4c
ストライプ 5	PB5	PB5	DB5a	DB5b	DB5c	DB5d
ストライプ 6	PB6'	DB6a	DB6b	DB6c	DB6d	PB6

【図 5】



【図 6】

	記憶装置 30a	記憶装置 30b	記憶装置 30c	記憶装置 30d	記憶装置 30e	記憶装置 30f
スライブ 1	DB1a	DB1b	DB1c	DB1d	PB1	(DB1b)
スライブ 2	DB2b	DB2c	DB2d	PB2	(DB2c)	DB2a
スライブ 3	DB3c	DB3d	PB3	(DB3d)	DB3a	DB3b
スライブ 4	DB4d	PB4	PB4'	DB4a	DB4b	DB4c
スライブ 5	PB5	PB5	DB5a	DB5b	DB5c	DB5d
スライブ 6	(DB6a)	DB6a	DB6b	DB6c	DB6d	PB6

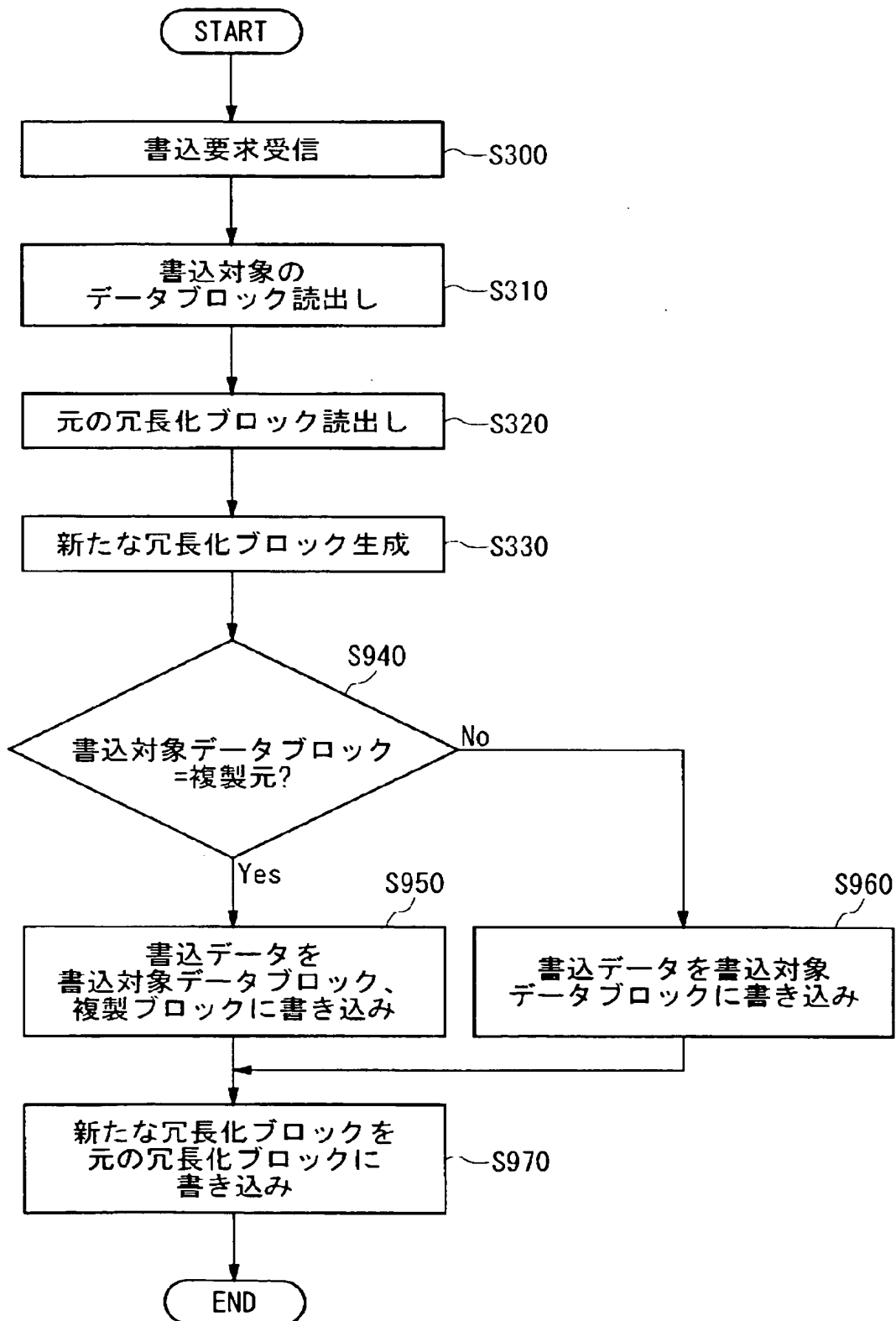
【図 7】

	記憶装置 30a	記憶装置 30b	記憶装置 30c	記憶装置 30d	記憶装置 30e	記憶装置 30f
ストライプ 1	DB1a	[DB1b]	DB1c	DB1d	PB1	(DB1b)
ストライプ 2	DB2b	[DB2c]	DB2d	PB2	(DB2c)	DB2a
ストライプ 3	DB3c	[DB3d]	PB3	(DB3d)	DB3a	DB3b
ストライプ 4	DB4d	[PB4]	PB4'	DB4a	DB4b	DB4c
ストライプ 5	PB5	[PB5']	DB5a	DB5b	DB5c	DB5d
ストライプ 6	(DB6a)	[DB6a]	DB6b	DB6c	DB6d	PB6

【図 8】

	記憶装置 30a	記憶装置 30b	記憶装置 30c	記憶装置 30d	記憶装置 30e	記憶装置 30f
ストライプ 1	DB1a	DB1b	DB1c	DB1d	PB1	DB1a'
ストライプ 2	DB2b	DB2c	DB2d	PB2	DB2a'	DB2a
ストライプ 3	DB3c	DB3d	PB3	DB3a'	DB3a	DB3b
ストライプ 4	DB4d	PB4	DB4a'	DB4a	DB4b	DB4c
ストライプ 5	PB5	DB5a'	DB5a	DB5b	DB5c	DB5d
ストライプ 6	DB6a'	DB6a	DB6b	DB6c	DB6d	PB6

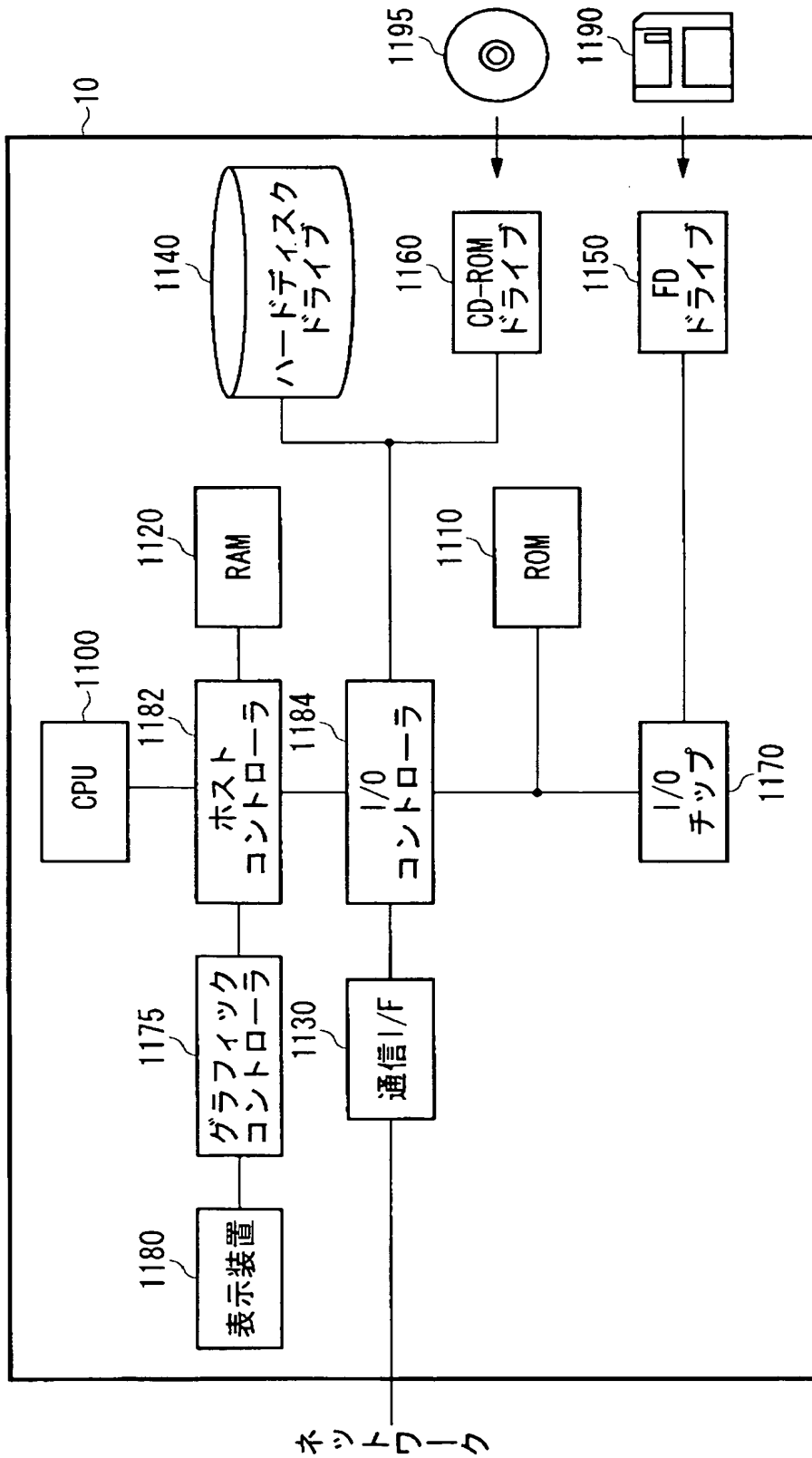
【図 9】



【図 10】

	記憶装置 30a	記憶装置 30b	記憶装置 30c	記憶装置 30d	記憶装置 30e	記憶装置 30f
ストライプ 1	DB1a	DB1b	DB1c	DB1d	PB1	(DB1b)
ストライプ 2	DB2b	DB2c	DB2d	PB2	(DB2c)	DB2a
ストライプ 3	DB3c	DB3d	PB3	(DB3d)	DB3a	DB3b
ストライプ 4	DB4d	PB4	(PB4)	DB4a	DB4b	DB4c
ストライプ 5	PB5	DB5a'	DB5a	DB5b	DB5c	DB5d
ストライプ 6	DB6a'	DB6a	DB6b	DB6c	DB6d	PB6

【図 11】



【書類名】 要約書

【要約】

【課題】 複数の格納対象ブロックからなるブロックグループを複数の記憶装置に分散して格納する記憶システムにおいて、一の記憶装置が故障した場合におけるブロック再生処理のオーバーヘッドを低減する。

【解決手段】 一の格納対象ブロックは、他の複数の格納対象ブロックを再生するための冗長化ブロックであり、複数の格納対象ブロックのそれぞれと、複数の格納対象ブロックのいずれかを複製した複製ブロックとを、互いに異なる記憶装置に格納するブロック書込部と、複製されていない格納対象ブロックの故障が検出された場合に、複数の格納対象ブロックのうち故障した格納対象ブロック以外のブロックに基づいて、故障した格納対象ブロックを再生するブロック再生部と、再生された格納対象ブロックを、複製ブロック又は複製ブロックの複製元となった格納対象ブロックに上書きする再生ブロック上書部とを備える記憶システムを提供する。

【選択図】 図 1

認定・付加情報

特許出願の番号	特願 2003-118907
受付番号	50300680236
書類名	特許願
担当官	末武 実 1912
作成日	平成 15 年 6 月 4 日

<認定情報・付加情報>

【特許出願人】

【識別番号】	390009531
【住所又は居所】	アメリカ合衆国 10504、ニューヨーク州 アーモンク ニュー オーチャード ロード
【氏名又は名称】	インターナショナル・ビジネス・マシーンズ・コーポレーション

【代理人】

【識別番号】	100086243
【住所又は居所】	神奈川県大和市下鶴間 1623 番地 14 日本アイ・ビー・エム株式会社 大和事業所内
【氏名又は名称】	坂口 博

【代理人】

【識別番号】	100091568
【住所又は居所】	神奈川県大和市下鶴間 1623 番地 14 日本アイ・ビー・エム株式会社 大和事業所内
【氏名又は名称】	市位 嘉宏

【復代理人】

【識別番号】	100104156
【住所又は居所】	東京都新宿区新宿 1 丁目 24 番 12 号 東信ビル 6 階 龍華国際特許事務所
【氏名又は名称】	龍華 明裕

【代理人】

【識別番号】	100108501
【住所又は居所】	神奈川県大和市下鶴間 1623 番 14 日本アイ・ビー・エム株式会社 知的所有権
【氏名又は名称】	上野 剛史

次頁無

特願 2003-118907

出 願 人 履 歴 情 報

識別番号

[390009531]

1. 変更年月日

2000年 5月16日

[変更理由]

名称変更

住 所

アメリカ合衆国10504、ニューヨーク州 アーモンク (番地なし)

氏 名

インターナショナル・ビジネス・マシーンズ・コーポレーション

2. 変更年月日

2002年 6月 3日

[変更理由]

住所変更

住 所

アメリカ合衆国10504、ニューヨーク州 アーモンク ニュー オーチャード ロード

氏 名

インターナショナル・ビジネス・マシーンズ・コーポレーション